

Vidjil, une alliance entre informatique et biologie pour mieux suivre la leucémie

Depuis 2011, Mathieu Giraud et Mikaël Salson, de l'équipe-projet Bonsai d'Inria Lille – Nord Europe (commune avec le CNRS et l'université Lille 1*), développent Vidjil, un logiciel open-source capable de mettre en forme l'analyse de millions de données issues des lymphocytes de personnes atteintes de leucémie. L'objectif est de suivre l'évolution de la maladie pour, à terme, ajuster le traitement en conséquence. Une collaboration entre les bioinformaticiens et des biologistes a, notamment, permis d'amorcer le transfert de leur technologie au centre de biologie pathologie du CHRU de Lille dès janvier 2015.

Vidjil : l'analyse de segments d'ADN nommés V, D et J

Lorsqu'un patient souffre de leucémie, une analyse biologique quantifie le nombre de globules blancs cancéreux. Or, mille milliards de lymphocytes circulent dans le sang d'un être humain, il faut repérer ceux atteints par la maladie. Chacun d'entre eux possède une séquence ADN propre. Plus précisément, leur diversité se caractérise par la variation de fragments, appelés V (gène variable), D (gène de diversité) ou J (gène de jonction), contenus dans l'ADN du lymphocyte. La combinaison de chaque V, D et J offre de nombreuses possibilités de séquences. Ainsi lorsque l'on possède la séquence ADN d'un lymphocyte – grâce à une méthode appelée séquençage –, il suffit d'analyser les combinaisons de V, D et J présentes pour différencier les lymphocytes. La fluctuation du nombre de globules blancs possédant une même combinaison est un indicateur de l'évolution de la pathologie. Et c'est sur cette caractéristique que s'appuie le logiciel lillois Vidjil pour présenter au médecin un suivi de la leucémie, cancer des lymphocytes conduisant à une prolifération anormale de globules blancs possédant une certaine combinaison V, D et J (figure 1). « *Vidjil repère les combinaisons V, D et J et propose une quantification de chaque clone, c'est-à-dire chaque type de lymphocytes, dont les cancéreux, au fil du temps* », révèle Mikaël Salson, maître de conférence à l'université Lille 1.

GGAAGGGCAGAATTA
V4 GGATGGG GAATTA J2

Figure 1 - Séquence ADN du lymphocyte V4J2 - Ce lymphocyte présente les segments des gènes V4 et J2 dans sa séquence ADN. Le nucléotide T du gène V4 est muté en A dans la séquence du lymphocyte © Bonsai

D'une à des millions : le séquençage haut débit

Jusqu'à maintenant, les biologistes séquençaient les lymphocytes uniquement grâce à la méthode Sanger. Ils repéreraient et quantifiaient une seule population de cellules cancéreuses. « *Sauf que d'autres types de lymphocytes cancéreux pouvaient se développer en parallèle, ou même le lymphocyte initial pouvait muter, c'est à dire modifier sa séquence et passer inaperçu* », explique Mikaël Salson. Ainsi, les résultats d'analyses ne pouvaient percevoir certaines rechutes. La solution est d'observer l'ensemble des lymphocytes et donc de gérer des millions de données. Ceci est devenu possible grâce à l'avancée des technologies. « *Tous les dix-huit mois, les ordinateurs vont deux fois plus vite, effectuent deux fois plus de calculs. Là, c'est encore plus rapide, presque tous les ans, on peut séquençer dix fois plus de séquences* », explique Mathieu Giraud, chercheur CNRS. Mais il fallait aussi que la méthode de séquençage évolue en elle-même. Mikaël Salson raconte que « *depuis 2004, les puces contiennent à leur surface des millions de trous, des petits séquenceurs où chaque séquence peut être analysée simultanément* ». Ainsi, Vidjil, le logiciel conçu par les bioinformaticiens, récupère les millions de données issues de séquenceurs dit à haut débit. Il organise les séquences des lymphocytes en clones et calcule les concentrations relatives à chacun. Enfin ce n'est que la partie cachée. La partie visible du logiciel, notamment pour les hématologues, ressemble à une application web où plusieurs courbes expriment le taux de tel clone dans le sang en fonction du temps, et donc de l'évolution de la maladie (figure 2).

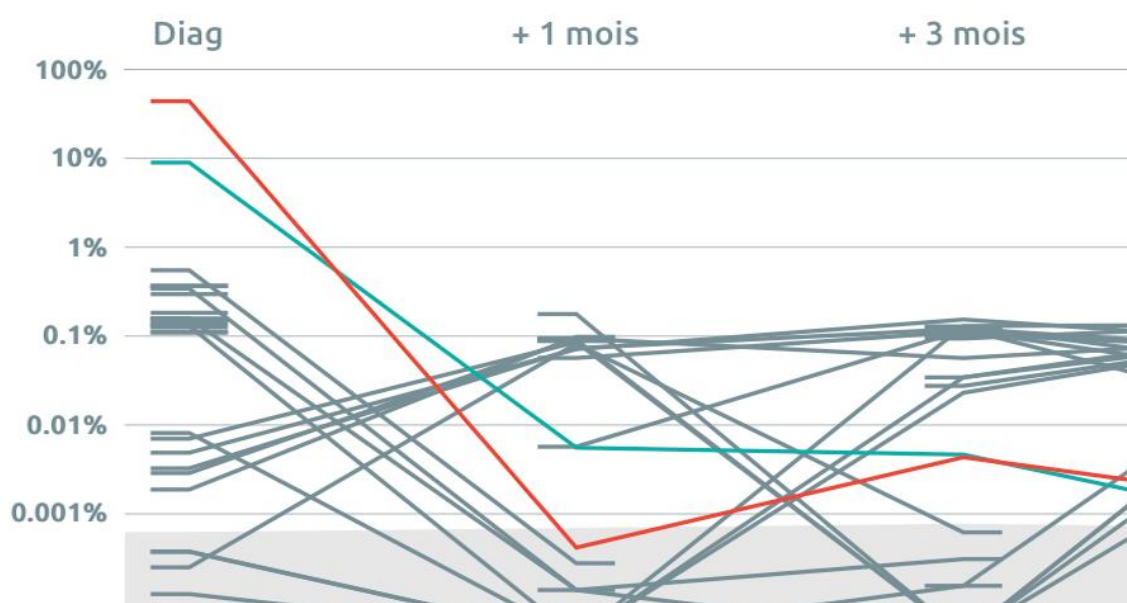


Figure 2 - Visualisation d'un résultat d'analyse via Vidjil - Seuls les clones ayant les concentrations les plus élevées sont représentés. Toutefois en interne, Vidjil a analysé les séquences de l'ensemble des globules blancs du prélèvement. Le clone représenté en rouge est aux alentours de 40 % au diagnostic et à moins de 0.001 % au bout d'un mois, après la phase initiale de la thérapie, l'induction. © Bonsai

Vidjil en cours de transfert au CHRU de Lille

L'avancée technologique n'a pas que permis d'aller plus vite. Le séquençage haut débit est aussi moins coûteux avec le temps. Les chercheurs révèlent qu'« *aujourd'hui les séquenceurs haut débit sont présents partout en France, et même, partout dans le monde.* » Le coût est si bas que ce séquençage devient accessible aux hôpitaux dans leur quotidien. « *Dans quelques*

années, le médecin proposera un séquençage d'un génome entier au même titre qu'une prise de sang», éclaire Mathieu Giraud. D'ailleurs, Vidjil a investi les machines d'hôpitaux français et européens dont le centre de biologie pathologie de Lille depuis le 1^{er} janvier 2015. À Lille, les lymphocytes de 36 patients diagnostiqués avec une leucémie ont été analysés via le logiciel Vidjil. « Mais aussi avec la méthode habituelle en parallèle », ajoute Mathieu Giraud, avant d'expliquer : « On ne change pas les protocoles hospitaliers en claquant des doigts, et heureusement ! Il existe un consortium fixant les 'règles'. Avec les machines assez récentes, il y a une mise au point à faire du côté technique et informatique. Il faut avoir du recul. Et ça va prendre du temps, avec une échéance de cinq à dix ans. » D'ici là, le logiciel open source continue son expansion et son développement informatique. « En dehors du suivi des leucémies que nous commençons à bien connaître, nous aimerions trouver des collaborations, que cela soit avec des laboratoires ou des industries, dans le domaine de l'immunologie. En effet, une application de Vidjil en immunologie plus générale, en étude de la réponse immunitaire ou des vaccins, est tout à fait envisageable », concluent les chercheurs.

*au sein de l'UMR 9189 CNRS - Centrale Lille - Université Lille 1, CRISTAL

Inria Lille - Nord Europe | Les sciences du numérique au cœur de l'Eurorégion

Inria, institut de recherche dédié au numérique, promeut « l'excellence scientifique au service du transfert technologique et de la société ». Inria emploie 2700 collaborateurs issus des meilleures universités mondiales, qui relèvent les défis des sciences informatiques et mathématiques. L'institut rayonne sur le territoire français à travers ses huit centres de recherche.

Le centre de recherche Inria Lille - Nord Europe, créé en 2008, compte 360 personnes dont 300 scientifiques répartis dans seize équipes de recherche. Reconnu pour son implication forte dans le développement socio-économique sur le territoire du Nord - Pas-de-Calais, le centre de recherche Inria Lille - Nord Europe s'engage, aux côtés de ses partenaires académiques, institutionnels et industriels, sur les grands chantiers de recherche et d'innovation dans le domaine du numérique.

La stratégie du centre est de développer autour de la métropole lilloise un pôle d'excellence de rayonnement international et à fort impact local. Pour ce faire, elle s'appuie sur des thématiques de recherche ambitieuses dans le domaine des sciences du numérique : l'intelligence des données et les systèmes logiciels adaptatifs. En favorisant les synergies entre chercheurs et industriels, Inria participe au transfert de compétences pour les innovations technologiques du numérique.

Rédaction : Nadège Joly | presselille@inria.fr